American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)

ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Machine Vision Detection Method for Surface Defects of Automobile Stamping Parts

Lei Geng[a], YuXiang Wen[b], Fang Zhang[c]*, YanBei Liu[d]

[a,b,c,d]*School of Electronic and Information Engineering, Tianjin Polytechnic University, Tianjin 300387, China*

[a]*Email: genglei@tjpu.edu.cn*

[b]*Email: dr-wyx@qq.com*

[c]*Email: hhzhangfang@126.com*

[d]*Email: liuyanbei@tjpu.edu.cn*

Abstract

A new system for detecting surface defects of automobile stamping parts is designed. Set up a defect detection system, which includes light source, camera and lens. Through this system, auto body surface image is collected, defect data set is established and manual annotation is performed. The marked training set was used to train the convolutional neural network, and then the convolutional neural network was used to segment the surface defects at the pixel level to obtain the defect regions of different categories. The experimental results show that the surface defect detection method proposed in this paper can effectively segment the surface defects of the workpiece, and the accuracy of scratch and rust segmentation is high.

Keywords: Surface defect detection; Machine vision; Convolutional neural network; Camera calibration; Image segmentation.

1. Introduction

Automobile stamping is one of the most important components of automobile, which directly reflects the overall quality and aesthetics of automobile. As the most common defects of body parts, scratches and rust are important factors affecting the quality of body parts. If these defects are not accurately detected, it will have a great impact on the subsequent process. Therefore, it is of great significance to detect and process the surface defects of automobile body parts timely and accurately. Surface defect detection refers to the detection of defects on the surface of the workpiece under test by using thermal, acoustic, optical, electrical, magnetic and other characteristics without affecting the working performance of the workpiece or material under test.

---

* Corresponding author.

The defect detection method has gone through the stages of combining manual visual inspection with later sampling inspection, single electromechanical or optical technology, and integrated photoelectric machine vision detection[1]. The combination of manual visual inspection and later sampling inspection requires low cost and no special equipment is needed. However, this method has many disadvantages, such as high rate of omission, low detection efficiency and high labor cost. With the development of science and technology, surface defect detection technology has entered into a single electromechanical or optical technology detection stage. These techniques can be called nondestructive testing techniques. At present, the most widely used nondestructive testing technologies include magnetic particle testing, eddy current testing, penetration testing, ultrasonic testing and machine vision testing. With the rapid development of machine vision technology, integrated photoelectric detection technology has been applied in various aspects of industrial production, such as machinery, electronics, printing, textile and other industries. Machine vision detection is a typical representative of the photoelectric integration detection stage. Machine vision detection technology is to obtain the surface image of the workpiece by using the camera instead of human eyes, and then judge whether there are defects on the surface of the workpiece by processing the image through the computer. A complete set of machine vision system includes: industrial camera, light source, photoelectric sensor, image acquisition card, computer, etc. The image of the detected workpiece is collected by an industrial camera and a suitable lens for the application. The camera will collect the image through the image acquisition card transfer to the computer memory. The machine vision software processes the collected images and gives the detection results. In this paper, automotive stamping parts are selected as the research material to detect the surface defects. Magnetic particle testing technology, ultrasonic testing technology requires the medium coating on the material surface, causing wear to the material surface. The surface wear of the stamping parts after spray painting causes diffuse reflection of light, which makes the wear visible. It not only affects the appearance of the stamping parts, but also causes high repair cost. Eddy current detection technology is not suitable for industrial robots to quickly detect defects due to its difficulty in implementation. Machine vision detection technology has the advantages of simple structure, high sensitivity and fast detection speed, so this paper adopts this detection technology for relevant research. Although the surface defect detection technology based on machine vision has many advantages, has obtained many new achievements in theoretical research, has been widely applied in industrial practice and achieved remarkable detection effect, there are still many problems and difficulties:

The surface defect detection technology based on machine vision is susceptible to light, noise and other factors in the field environment. The changes of these factors bring great difficulties to the defect detection algorithm, so how to build a stable defect detection system to weaken the interference caused by environmental factors has become the primary problem to be solved.

There are many types and shapes of surface defects of automobile stamping parts, and the image background acquired by the visual defect detection system is complex. These factors make it difficult to define the image features of defects, which bring great difficulty to defect extraction and leads to low defect recognition rate.

In industrial processing sites, defect detection is often carried out by online detection, which is characterized by large amount of data, rich information and high spatial dimension. Therefore, it is necessary to ensure the real-time performance of defect detection while accurately obtaining the required defect information in a large

amount of data.

Although many excellent defect detection algorithms based on machine vision have been proposed by domestic and foreign experts and scholars, there is still a certain gap between the accuracy of defect detection in practical applications and the field requirements[2-3]. How to improve the accuracy of defect detection is still a difficult problem in current research.

For the above problems, this paper designs a defect detection system based on vision sensor. The hardware part of the system mainly includes the robot defect detection system and the workpiece to be tested, among which the robot defect detection system includes industrial robot, industrial camera and annular light source, and the software part is mainly defect detection software. By selecting the appropriate hardware model, the interference of the surrounding environment to the defect detection is reduced and the difficulty of the defect detection algorithm is reduced. By designing an appropriate defect detection algorithm, the accuracy and speed of defect detection are guaranteed.

2. Methods

In this section, we mainly introduce the visual sensor-based defect detection system and system calibration algorithm. In order to detect missing test accurately, a new convolutional neural network is designed based on vgg-16 network.

2.1 Defect detection system design

The hardware of visual sensor-based defect detection system mainly includes industrial camera, lens and light source. The selection of the appropriate hardware model is crucial for defect detection. The construction of a suitable visual acquisition system can greatly reduce the complexity of the algorithm, and the quality of the image acquired by the camera will directly affect the accuracy of defect detection. The defect detection system based on vision sensor needs to cooperate with industrial robots, and the camera performance, the speed of defect detection algorithm and the moving speed of robot body are closely related. In order to ensure the accuracy of defect detection, the accuracy of the image collected by the camera has become the primary factor. Improving the image accuracy means improving the resolution of the camera. However, improving the resolution of the camera also means that the frame rate of the camera acquisition and the speed of the defect detection algorithm decrease. Therefore, the selection of suitable hardware becomes the basis of the defect detection system.

Industrial camera is the core of vision sensor. The selection of industrial camera directly affects the accuracy of subsequent algorithms in dealing with defects. There are four main factors affecting the performance of the camera:

Output signal type. Industrial cameras can be divided into analog and digital cameras according to the types of output signals. Compared with analog camera, digital camera has the advantage of strong anti-interference ability.

Chip type. The most important component of industrial camera is sensor chip, which mainly includes charge coupled device (CCD) and complementary metal oxide semiconductor (CMOS).CCD sensor has the advantages of high imaging quality and low noise compared with CMOS sensor.

Resolution. The higher the resolution of an industrial camera, the more pixels in the image it captures, and the more detailed the features of the object.

Frame rate. The higher the camera's frame rate, the more images the camera can capture per unit time. The frame rate of camera directly affects the speed of defect detection.

In the visual sensor-based defect detection system designed in this paper, it is necessary to carry out defect detection on the surface of car body parts with obvious surface defect characteristics. Therefore, more attention should be paid to the resolution of the camera than the frame rate. In this paper, the design size of the field of view for defect detection is 100 75mm2, and the theoretical accuracy is 0.05mm, so the single-direction resolution of the camera is 2000pix and 1600pix, respectively, and the resolution of the camera is at least 3.2 million. Based on the above considerations, this paper selects the mv-ce050-50gm industrial camera produced by hikvision, as shown in Figure 1. The camera has a resolution of 5 million and a frame rate of 14fps, which meets the design requirements of the system. The specific parameters are shown in Table 1.

Figure 1: MV- CE050-50GM camera

Table 1: MV- CE050-50GM camera parameters

Index	Parameter
Resolution (pix)	2592×1944
Pixel Size(um)	2.2×2.2
Dynamic Range(dB)	>60
Frame Rate(fps)	14
Lens Interface	C-Mount
Power	12 VDC

Temperature(°C)	0~50

Lens is an optical device used to focus light on a digital sensor to create an image. The main parameters that affect lens selection include: focal length, aperture, field of view and target surface size. In defect detection, the distance between the camera lens and the object is relatively close, so the lens with a large focal length can be selected. In this paper, according to object distance, CCD target surface size and field of view size, the focal length of the optical lens shall be calculated by formula 1.

$$f = w\frac{H}{D} = 5.76 \times \frac{200}{100} = 11.52mm \quad (1)$$

In the above equation, D is the object distance and the size is 200mm; w represents the size of the selected CCD target surface, with a size of 5.76mm. H is the length of field of view, with a size of 100mm. V is the width of field of view and the size is 80mm.Then the required perspective is:

$$\theta_H = 2\arctan\frac{H}{2D} \approx 28° \quad (2)$$

$$\theta_V = 2\arctan\frac{W}{2D} \approx 23° \quad (3)$$

To sum up, MVL-MF1220M -5MP lens is selected as the lens of the measurement system in this paper, as shown in Figure 2, and the detailed parameters are shown in Table 2.

Figure 2: MVL-MF1220M -5MP lens

Table 2: MVL-MF1220M -5MP camera parameters

Index	Parameter
Focal Length (mm)	12.00
Aperture	F2.0 - F16.0
Image Circle	2/3"
M.O.D	80 mm
Field of View(°)	H:40.2 V:30.6

Temperature(°C)	-10～50

The purpose of illumination in machine vision is to make the important features of the object to be measured appear. Light source as one of the indispensable equipment in the machine vision systems, direct affect the imaging quality of the camera, the right lighting design can not only highlight the important feature of the object to be tested, but can be the good characteristics of don't need to contain, reduce interference, reduce the difficulty of subsequent algorithm processing, improve the robustness of the whole system. According to the characteristics of the workpiece to be tested, a low-angle circular light source is used in this paper to polish the workpiece directly above, as shown in Figure 3.

Figure 3: Circular light source

2.2 Visual system calibration

The images collected by the visual system on the surface of automobile body parts are two-dimensional images. The location information of defects detected by the defect detection algorithm is also two-dimensional image coordinate information. In order to map the defect image position to the position in the real space, this paper adopts the camera perspective projection model, and the process of obtaining the mapping relationship by solving the model parameters is camera calibration[4-5].

2.2.1 Camera perspective projection model

Camera imaging model is divided into linear model and nonlinear model. Without considering the lens distortion, the camera imaging model is a linear model. The perspective projection model is the most commonly used linear model, as shown in Figure 4. Perspective projection model is to project the three-dimensional model of the real world onto the two-dimensional plane of the camera lens. This process can be represented by the coordinate transformation between the four coordinate systems.

The perspective projection model consists of four coordinate systems: camera coordinate system $O_C - X_C Y_C Z_C$, world coordinate system $O_W - X_W Y_W Z_W$, image plane coordinate system $o' - xy$ and image coordinate system $o - uv$. The world coordinate system and the camera coordinate system are coordinates in the object space, the image plane coordinate system and the image coordinate system are coordinates in the image space. The coordinate of 3d point in the world coordinate system is (X_W, Y_W, Z_W), the coordinate of camera coordinate

system is (X_C, Y_C, Z_C), the coordinate of image point in the image plane coordinate system is (x, y), and (u, v) is the corresponding position in the image coordinate system.

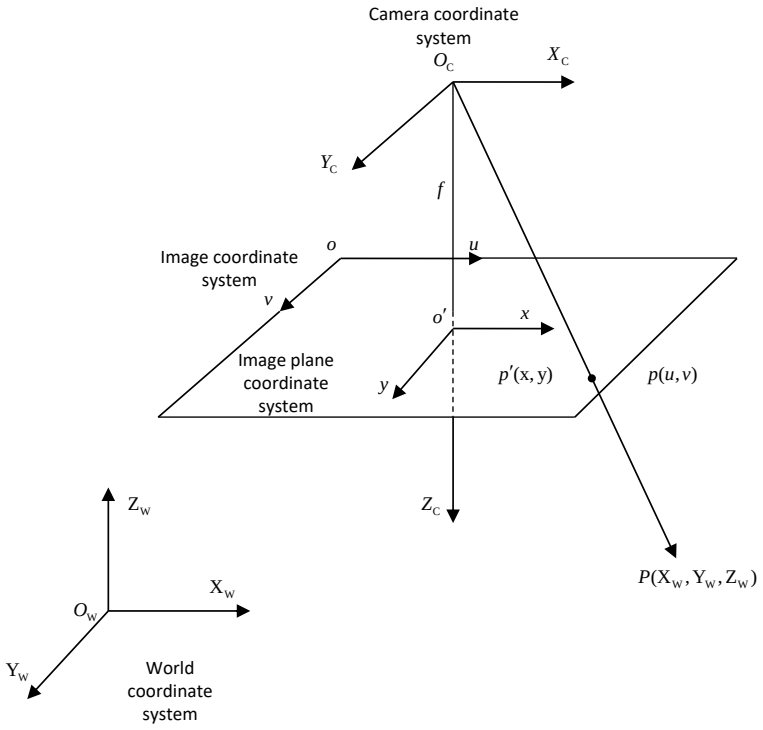


Figure 4: Perspective projection model

- World coordinate system $O_W - X_W Y_W Z_W$

The world coordinate system is also called the absolute coordinate system. It is the three-dimensional coordinate system of the actual space between the camera and the object under test. The coordinate system can be established in a certain position in the real space according to the actual needs.

- Camera coordinate system $O_C - X_C Y_C Z_C$

The origin O_C of the camera coordinate system is the camera optical center, and the z-axis coincides with the camera optical axis, which is perpendicular to the image plane and takes the shooting direction as the positive direction. The axes X_C and Y_C are parallel to the axes x and y in the plane coordinate system.

- Image plane coordinate system $o' - xy$

The image plane coordinate system is based on the camera CCD chip plane. The origin of the coordinate system is the intersection of the optical axis and the imaging plane. The coordinate of the origin in the camera coordinate system is $(0, 0, f)$. As can be seen, the relationship between the plane coordinate system and the camera coordinate system is the translation along the z axis. The translation is the focal length of

the camera, and the coordinate measurement unit in the imaging coordinate system is mm.

- Image coordinate system $o-uv$

The CCD sensor chip of the camera inputs the collected optical signals into the computer in the form of standard electrical signals. The collected signals are converted into digital images that can be processed by the image acquisition card. Each digital image can be viewed as an M*N matrix and each value in the matrix corresponds to the gray value of each pixel. The coordinate system defines the origin at the upper left corner of the entire pixel matrix, and the unit of measurement is pixels.

The transformation from the world coordinate system to the camera coordinate system is a rigid transformation composed of translation and rotation. The transformation from camera coordinate system to image physical coordinate system is transformed into perspective transformation. Finally, the transformation between image plane coordinate system and image coordinate system is completed through the affine transformation of coordinate scaling and translation. The mathematical derivation of the transformation between each coordinate system is as follows:

The transformation from the world coordinate system to the camera coordinate system

Since the world coordinate system and the camera coordinate system belong to the rigid transformation, the world coordinate system can be transformed into the camera coordinate system after translation and rotation. The mathematical expression of its transformation relation is as follows:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{4}$$

The transformation relation between camera coordinate system and image physical coordinate system

Perspective projection is the process of transforming the coordinate points in the three-dimensional space from the camera coordinate system to the two-dimensional plane coordinate system of the camera imaging. In order to align the image physical coordinate system with the image physical coordinate system (row coordinate increasing downward, column coordinate increasing to the right) to simplify the calculation process, the image plane is placed at A focal length f from the center of the camera projection. In Figure 4, point P in the camera coordinate system is converted to point P' in the object imaging plane coordinate system, and according to the triangle similarity principle:

$$Z_C \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \tag{5}$$

The transformation relation between image physical coordinate system and image pixel coordinate system

The object image captured by the camera can be represented as a two-dimensional array containing M rows and N columns. The element value of each position in the array is the grayscale value of the digital image at that point. In the image coordinate system u_i and v_i, respectively, represent the number of columns and rows where the pixel points are. Point (u_i, v_i) is the main point of the digital image, that is, the representation of the vertical projection point of the optical center of the camera lens in the image plane in the image pixel coordinate system. d_x、d_y are the distances between adjacent pixels in the direction of u and v respectively. Then the points in the pixel coordinate system of the image can be expressed as follows:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \begin{bmatrix} \frac{x}{d_x} + u_o \\ \frac{y}{d_y} + v_o \end{bmatrix} \tag{6}$$

Since then, the conversion relation between adjacent coordinate systems in Figure 4 has been deduced, and the conversion formula from the world coordinate system to the image pixel coordinate system is obtained by formula (4), (5) and (6) as follows:

$$\begin{aligned} Z_c \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{s_x} & 0 & u_o \\ 0 & \frac{1}{s_y} & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f_x & 0 & u_o & 0 \\ 0 & f_y & v_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = M_1 M_2 P_w = M P_w \end{aligned} \tag{7}$$

2.2.2 Distortion model

There is no distortion in the ideal image coordinate system, and the camera imaging model is divided into linear model. However, in practice, due to the influence of camera lens processing technology and other factors, the lens is often distorted, resulting in different magnification of different areas of the image on the focal plane, resulting in image distortion, geometric distortion and other phenomena. Without distortion correction, the detection accuracy will decrease. There are two kinds of lens distortion in practice: radial distortion and tangential distortion.

Radial distortion

Radial distortion is a phenomenon that pixel points are offset in the radial direction. It is common in most

lenses, and the distortion coefficient is K. Radial distortion is symmetric about the main optical axis of camera lens, which is divided into pillow distortion and barrel distortion. Because the magnification of the lens edge is higher than the central position, the edge of the image is concave inward, so the pillow distortion is also known as negative distortion. Because the magnification of the lens edge is higher than the center position, the edge of the image bulges outward, so the pillow distortion is also known as the positive distortion. Assume that the ideal coordinate of a point in the image coordinate system in the camera coordinate system is (x,y), and the radial distortion value is $(\Delta x_r, \Delta y_r)$, then the mathematical model of radial distortion is:

$$\begin{cases} \Delta x_r = x\left(K_1 r^2 + K_2 r^4 + K_3 r^6 + ...\right) \\ \Delta y_r = y\left(K_1 r^2 + K_2 r^4 + K_3 r^6 + ...\right) \end{cases} \tag{8}$$

$$r = \sqrt{x^2 + y^2} \tag{9}$$

Where, K_1、K_2 and K_3 are radial distortion coefficients.

Tangential distortion.

Tangential distortion is caused by the misalignment of optical lens' optical center and geometric center. Assuming that the ideal coordinate of A point in the image coordinate system in the camera coordinate system is (x,y) and the tangential distortion value is $(\Delta x_d, \Delta y_d)$, then the mathematical model of tangential distortion is:

$$\begin{cases} \Delta x_d = 2P_1 xy + P_2\left(r^2 + 2x^2\right) \\ \Delta y_d = P_1\left(r^2 + 2y^2\right) + 2P_2 xy \end{cases} \tag{10}$$

Where, P_1 and P_2 are tangential distortion coefficients.

2.3 Network design

Since it was first proposed in 2015, the full convolutional neural network has been widely applied due to its excellent performance in image segmentation [6]. By taking advantage of the spatial redundancy of adjacent pixels through convolution operation, the full connection layer in the network is replaced by convolution layer, which improves the computational efficiency to a certain extent. However, the full convolutional neural network does not have statistical efficiency, because the spatial redundancy brought by convolution operation limits the learning of information between adjacent pixels [7]. The network designed in this paper adopts the pixel hierarchical sampling strategy to increase the sample diversity in the process of batch stochastic gradient descent updating, and at the same time introduces a complex nonlinear predictor to improve the classification accuracy of pixels. Its network structure is shown in

Figure 5.

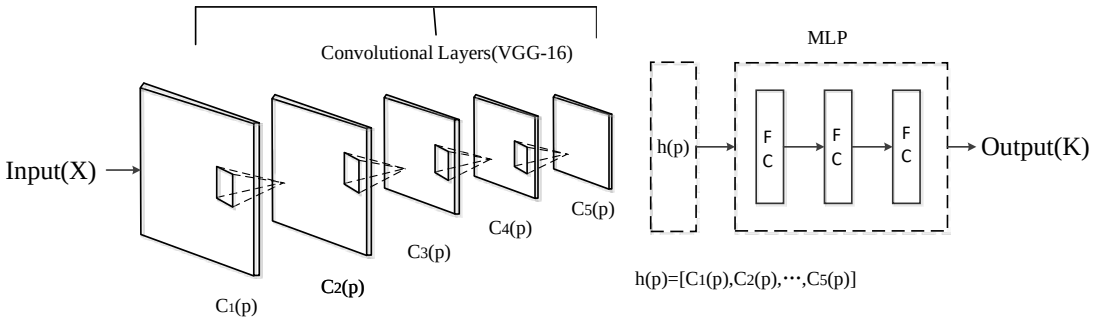


Figure 5: Network structure diagram

As shown in Figure 5, this network extracts multi-scale features behind the vgg-16 basic network. In the convolutional neural network, the deeper the convolutional layer is, the larger the corresponding receptive field will be. Therefore, higher level of image global context information can be obtained, but low-level image details may be lost at the same time. In this paper, the network establishes a predictor on the multi-scale features extracted by the multi-layer convolutional neural network, and uses the multi-scale hypercolumn descriptor to refer to the features extracted by multiple layers corresponding to the same pixel. The multi-scale hypercolumn features of pixel P are calculated as follows:

$$h_p(X) = [c_1(p), c_2(p), \cdots, c_M(p)] \qquad (11)$$

Where, $c_i(p)$ is the i layer convolution with pixel P as the center and X is the input image. The network transforms the image level pixel prediction problem into hypercolumn feature. The final prediction of pixel P can be given by the following formula:

$$f_{\theta,p}(X) = g(h_p(X)) \qquad (12)$$

3. Experiment and results analysis

Aiming at the design of the defect detection system and algorithm in this paper, experiments are designed to verify and evaluate the effectiveness of the system and algorithm. The experiment evaluates the performance of the algorithm through different indexes and drew a table to assist the description.

3.1 The experiments of visual system calibration

In camera calibration, it is necessary to obtain enough feature points, and feature points are required to be easy to extract to obtain their pixel coordinates. Internal and external camera parameters can be solved according to the transformation relation between the actual coordinates of feature points on the calibration board and pixel coordinates of the calibration board image. The calibration template used in this paper is

composed of three small dots with a diameter of 3.5mm, and its machining accuracy is 0.001mm.In the whole field of view of the camera, the calibration board was moved and rotated many times to obtain 18 calibration board images at different positions and angles, as shown in Figure 6. The location information of the mark points is obtained by image processing method, and then the camera internal and external parameters and camera distortion coefficient are solved.

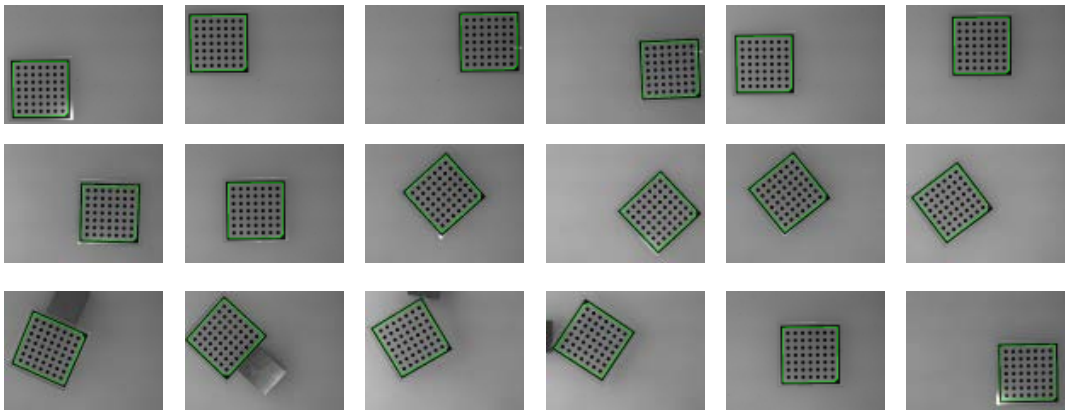


Figure 6: Calibration board images under different postures

The internal parameters of the camera obtained through experiments are shown in Table 3, and the distortion parameters are shown in Table 4.

Table 3: Calibration results of camera internal parameters

f_x	f_y	u_0 (pix)	v_0 (pix)
5512.5910	5517.6070	1293.6581	967.9156

Table 4: Calibration results of distortion parameters

k_1(1/m2)	k_2(1/m2)	k_3(1/m2)	p_1(1/m2)	p_2(1/m2)
-0.423×10^{3}	0.527×10^{6}	0.329×10^{9}	-0.145×10^{-2}	-0.298×10^{-1}

In order to test the calibration of the method, the camera calibration accuracy is evaluated by the re-projection error. The calculation formula is as follows:

$$\sigma = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(x_i' - x_i\right)^2 + \left(y_i' - y_i\right)^2} \qquad (13)$$

Where, $m_i = (x_i, y_i)$ is the coordinate of the feature point on the calibration image in the image coordinate system, $m_i' = (x_i', y_i')$ is its coordinate on the corresponding re-projection image, and N is the number of feature points on the calibration plate. The smaller the value of σ in the experiment, the higher the calibration accuracy. The re-projection error of this calibration experiment is 0.032017 pixels, which fully meets the detection requirements.

3.2 The experiments of defect detection

This paper takes automobile body parts as the research object, and the common surface defects are scratches and rust, as shown in Figure 7.Due to many surface textures of workpiece and different depth and shape of defects, the traditional segmentation algorithm cannot segment defects effectively. In view of the above problems, this paper uses convolutional neural network to complete the segmentation of workpiece surface defects.

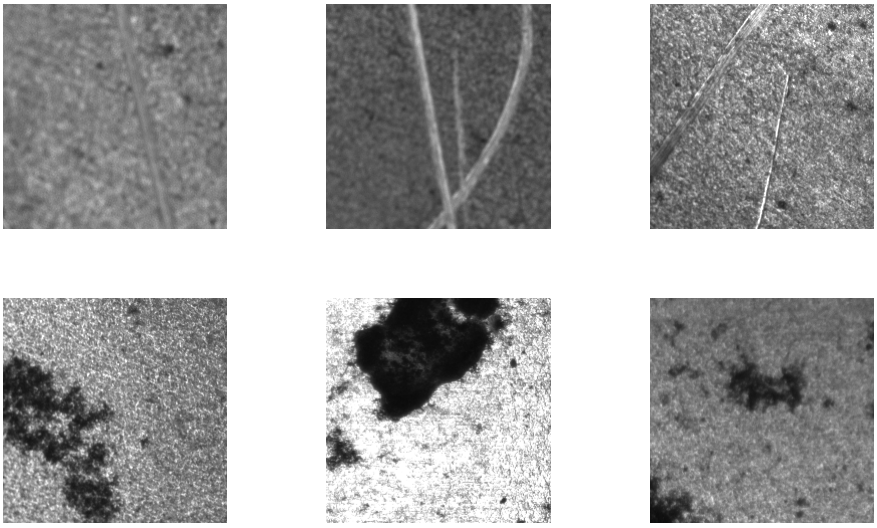


Figure 7: Scratches and rust

Deep convolutional network has strong expressive ability and requires a large amount of data to drive model training, otherwise overfitting will occur [8-9]. In practice, data expansion becomes an important step of deep model training. Effective data expansion can not only expand the number of training samples, but also increase the diversity of training samples. On the one hand, overfitting can be avoided; on the other hand, model performance can be improved. In this paper, a total of 3000 images were collected as samples, and the sample set was marked. Then the way of collecting surface defects of workpiece by visual sensor is simulated and the appropriate way of data expansion is selected. The pose of different samples at the same height is simulated by turning over and rotating the camera. The pose of the sample collected by the camera at different heights is simulated by scaling. The above two methods increase the robustness of the convolutional neural network to the scale and direction of the object. By adjusting the brightness of samples, the influence of field lighting environment on samples was simulated. Through the use of superposition of the above methods and the exclusion of some data, the final image data was expanded to

15,000.

In this model training, the Matlab development interface of deep learning framework Caffe is used in the Windows 7 operating system, and a single GPU is used for training. The specific hardware and software configuration information is shown in Table 4 and 5

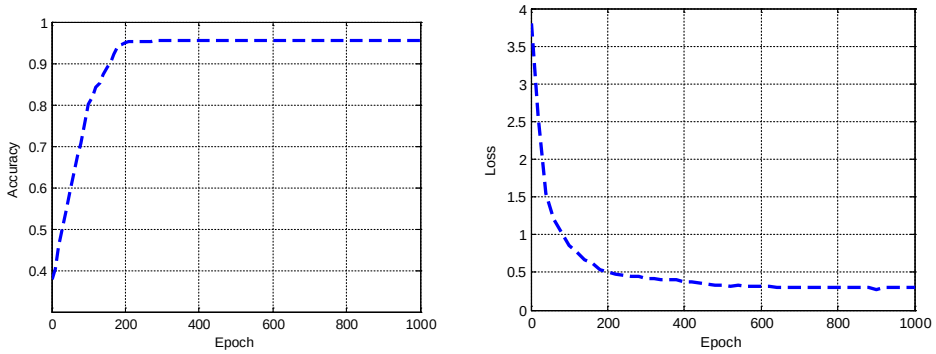
Table 4: Hardware configuration

Hardware	Model	Parameter
CPU	Intel Xeon E5-2625	2.1GHz
RAM	DDR3	32G
GPU	NVIDIA GTX1080Ti	11G
Hard disk	Mechanical drive	2T

Table 5: Software configuration

Software	Parameter
Operating system	Windows 7
Deep learning framework	Caffe
Cuda version	Cuda8.0
Developing interface	Matlab R2014A

9000 images were selected from the data set as the training set, 3000 as the verification set, and 3000 as the test set to train and test network. In the training process, 90 graphs were used to represent the step length of an iteration, in which the initial learning rate was set at 0.01. With the increase of the number of training rounds, it eventually decayed to 0.0001, and the number of iterations was 100,000.Changes in accuracy and loss during training are shown in Figure 8. The segmentation results of scratches and rust are shown in Figure 9.



(a)Accuracy change curve (b) Loss value change curve

Figure 8: Training process parameter change curve

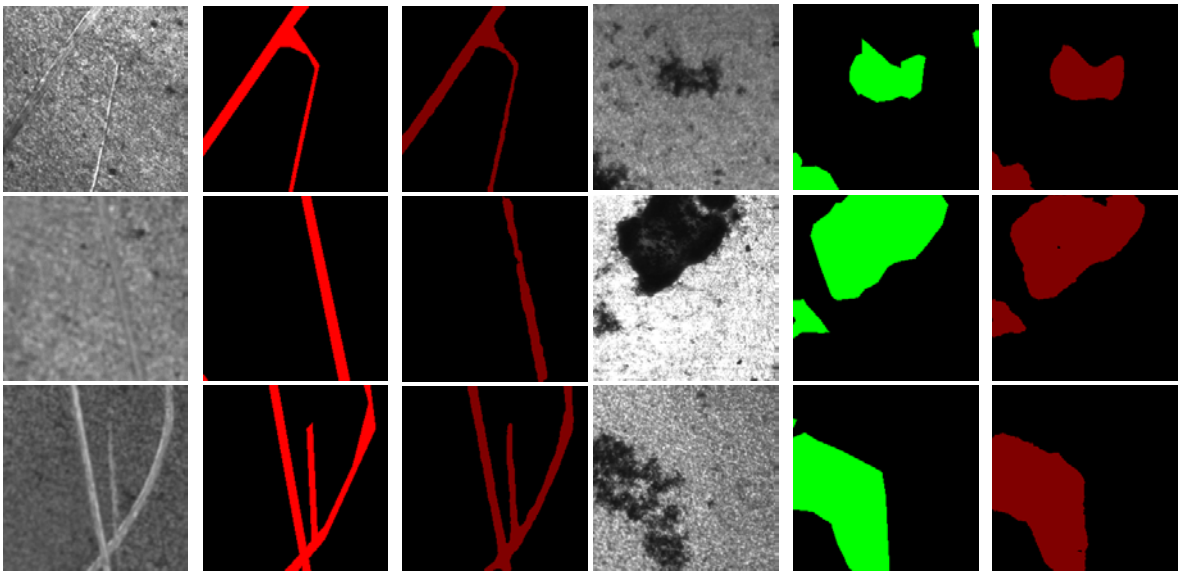


Figure 9: The segmentation results of scratches and rust

In order to objectively evaluate the segmentation effect of scratches with corrosion, this paper uses IOU and Acc [10-11] to evaluate the segmentation performance of the network on workpieces, as shown in Table 6. IOU refers to the intersection of two sets and set the ratio of the two set of real value respectively and predicted value, Acc is refers to the correct classification of the pixel number and ratio of all the pixel number, computation formula is as follows:

$$IOU = \frac{predicted\ \ segmentation \bigcap GroundTruth}{predicted\ \ segmentation \bigcup GroundTruth} \tag{14}$$

$$Acc = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \tag{15}$$

Table 6: Results of scratch and rust detection

Defect	Mean IOU	Mean Acc
scratch	0.9621	0.9985
rust	0.9579	0.9841

It can be seen from the defect segmentation results in Figure 9 that the defect segmentation method based on convolutional neural network adopted in this paper can complete the accurate segmentation of the surface defects of stamping parts. When the ambient light changes, scratches and rust still have good segmentation effect, which proves that the detection algorithm has a high robustness. It can be seen from Table 6 that the Mean IOU values of scratches and rust are 96.21% and 95.79%, respectively, which have a high accuracy and meet the requirements of actual defect detection and application.

3.3 Results analysis

The experimental results show that the surface defect detection method proposed in this paper can effectively segment the surface defects of workpieces, and the accuracy of scratch and rust segmentation is 96.21% and 95.79%, respectively. Therefore, the defect detection algorithm designed in this paper is feasible to a large extent. The defect detection system designed in this paper can accurately realize the function of defect detection.

4. Conclusion

A new system for detecting surface defects of automobile stamping parts is designed in this paper. Firstly, we select the hardware and build a visual inspection system. Then, we calibrate the visual detection system. Finally, the defects are segmented at the pixel level by the convolutional neural network. In order to verify the effectiveness of the algorithm in this paper, we use the defect images collected by the visual detection system to test the algorithm. Experimental results show that the algorithm proposed in this paper .

References

[1] Tang B, Kong JY. Overview of machine vision surface defect detection [J]. Journal of Image and Graphics, 2017, 22(12): 1640-1663.

[2] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2007, 9(1): 62-66.

[3] Yao MH, Chen ZH. Surface defect detection based on deep active learning [J]. Computer Measurement and Control, 2018, 26(9).

[4] Qiu ML, Li Y. Overview of camera calibration in computer vision [J]. Journal of Automation, 2000, 26(1): 43-55.

[5] Hou Z, Wang J. Study on traditional camera calibration [J]. Information-an International Interdisciplinary Journal, 2012, 15(11): 4393-4398.

[6] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the 2015 Conference on Computer Vision and Pattern Recognition, Boston, MA, 2015: 3431-3440.

[7] He K, X Zhang, S Ren, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016.

[8] Masci J, Meier U, Cireşan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[M]//Artificial Neural Networks and Machine Learning–ICANN 2011.Springer Berlin Heidelberg, 2011: 52-59.

[9] Simonyan K, Zisserman A. very deep convolutional networks for large-scale image recognition[C]. Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014: 1-14.

[10] Boureau Y, Ponce J, Lecun Y, et al. A theoretical analysis of feature pooling in visual recognition[C]. International Conference on Machine Learning, Haifa, Israel, 2010: 111-118.

[11] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines[C]. International Conference on Machine Learning, Haifa, Israel, 2010: 807-814.